Title
Bidimensional electronic spectroscopy on indole in gas phase and in water from first principles

Authors
Nenov, Artur
Rivalta, Ivan
Mukamel, Shaul
et al.

Publication Date
2014-07-01

DOI
10.1016/j.comptc.2014.03.031

Peer reviewed

Computational and Theoretical Chemistry 1040–1041 (2014) 295–303

Contents lists available at ScienceDirect

Computational and Theoretical Chemistry

journal homepage: www.elsevier.com/locate/comptc

Bidimensional electronic spectroscopy on indole in gas phase and in water from first principles

Artur Nenov [a,*], Ivan Rivalta [b], Shaul Mukamel [c], Marco Garavelli [a,b,*]

[a] Dipartimento di Chimica "G. Ciamician", Università di Bologna, V. F. Selmi 2, 40126 Bologna, Italy
[b] Laboratoire de Chimie, Ecole Normale Supérieure de Lyon, 46, allée d'Italie, 69364 Lyon, France
[c] Department of Chemistry, University of California, Irvine, CA 92697-2025, USA

ARTICLE INFO

Article history:
Received 6 February 2014
Received in revised form 27 March 2014
Accepted 27 March 2014
Available online 4 April 2014

Keywords:
Multiconfigurational methods
Double coherence
Excited states
Polarization continuum model

ABSTRACT

The electronic transitions of indole, the aromatic UV chromophore of the amino acid tryptophan, are characterized by using state-of-the-art multiconfigurational methods in gas-phase and aqueous solution, revealing the electronic spectrum up to 10 eV. Bidimensional near-ultraviolet (2D-NUV) electronic spectra of indole are simulated using the sum-over-states approach, based on ab initio calculations, accounting for different experimental set-ups, including rephasing ($KI = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$), quasi-absorptive (PP) and double quantum coherence ($KIII = \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3$) signals, and both one-color (2D-NUV) and two-colors (2D-NUV/Vis) regimes. In order to obtain accurate energies of high lying excited states and reliable 2DES spectra, extravalence virtual orbitals have been included using the restricted active space technique. The 2D-NUV spectrum of indole shows off-diagonal signals due to the correlation of the GS → L_b and GS → L_a transitions, an indole "fingerprint" in the NUV region that differentiate it from other aromatic chromophores in proteins. Further indole-specific transitions are resolved in the whole Vis–NUV range. A background-free region bellow the ionization potential, which shows no absorption signals in the monomer, can be used to resolve charge transfer states in coupled chromophore aggregates with a two-color 2D-NUV/Vis experimental set-up. Fundamental information for design of the 2DES experiments of indole is provided, including possible experimental pulse configurations that improve spectral resolution, revealing anharmonicities and selecting transitions. The proposed 2DES experiments could provide unprecedented level of detail for tracking indole electronic transitions in proteins, laying the groundwork for the use of nonlinear ultrafast optical spectroscopy for the study of protein structure and dynamics in solution.

1. Introduction

Indole, the chromophore of the aromatic amino acid tryptophan, is the most intensively absorbing protein chromophore in the near-ultraviolet (NUV). Its infrequent occurrence in proteins (~1.5% [1]) and characteristic absorption makes it a potential candidate for visualizing side chain interactions and resolving local stabilization factors during protein folding [2]. Linear response techniques such as circular dichroism spectroscopy, which reveal transitions to bright singly excited states in chromophores, are commonly used for differentiating secondary protein structures [3]. However, high spectral congestion makes the interpretation of such spectra challenging. Bidimensional (2D) electronic spectroscopy techniques are becoming an attractive tool with greatly improved spectral and temporal resolution [4–7]. Elaborate sequences of coherent femtosecond pulses are used to correlate the wavelength of the probed transitions to the wavelength of the pump pulse, thereby adding a second dimension for resolving the third-order nonlinear system response to the field-matter interaction [8–10]. In addition, 2D spectra contain qualitative new information missed by one-dimensional spectra because they visualize characteristic transition bands to higher lying states and correlations between electronic transitions. Due to the congested visible (Vis)/UV spectrum of aromatic chromophores a vast number of singly and doubly excited states can be probed, provided they possess non-vanishing transition dipole moments out of the doorway states prepared by interaction of the system with the pump pulses. As a consequence, every chromophore has its

* Corresponding authors. Address: Dipartimento di Chimica "G. Ciamician", Università di Bologna, V. F. Selmi 2, 40126 Bologna, Italy (M. Garavelli). Tel.: +39 051 2099476.
E-mail addresses: artur.nenov@unibo.it (A. Nenov), marco.garavelli@unibo.it (M. Garavelli).

http://dx.doi.org/10.1016/j.comptc.2014.03.031
2210-271X/

characteristic spectral signature, a fingerprint that could be used to discriminate it spectroscopically in molecular aggregates.

Spectral congestion further makes analysis of electronic spectra a challenging task for which theoretical modeling becomes crucial. The accurate characterization of 2D electronic spectra requires state-of-the-art quantum–mechanical methods describing electronic correlations [11]. The requirement of simultaneous treatment of a vast number singly and doubly excited states implies the usage of the state averaged (SA) complete active space self consistent field (CASSCF) approach [12] with second order perturbative treatment of the dynamic correlation of each state (SS-CASPT2) [13], i.e. SA-CASSCF//SS-CASPT2.

Here we present high-level multiconfigurational calculations preformed for the indole monomer in gas-phase and aqueous solution that provide an accurate description of its excited states below 10 eV. The spectrum below the first ionization potential at 7.76 eV [14] is assigned and a nomenclature for electronic transitions is proposed. Characteristic features of the indole spectrum are compared with benzene and phenol, the chromophores of the aromatic amino acids phenyalanine and tyrosine. 2D pump-NUV probe-NUV and pump-NUV probe–Vis spectra are discussed. The present study suggests possible experimental setups for novel 2D electronic spectroscopy of proteins in solution.

2. Computational methods

The indole ground state geometry was optimized at the CASSCF level [12] under C_S-symmetry with the generally contracted ANO-L basis set [15] adopting the following contraction scheme: C,O/[4s3p2d] and H/[2s1p]. It is well established that excited states described with the general ANO basis set are contaminated by spurious Rydberg character that can affect excited state energies and properties like transition dipole moments [16,17]. To account for this the basis sets was augmented with a set of 8s, 8p and 8d uncontracted Rydberg-type basis functions positioned at the center of nuclear charge of the lowest two cationic states [18,19] in all excited state calculations. We point out that we have used a set of uncontracted Rydberg-type basis functions instead of the usually utilized 1s, 1p, 1d contracted Rydberg-type set [20] in order to properly describe Rydberg states up to 10 eV. The orbital exponents for the uncontracted Rydberg basis functions were obtained from [21]. The uneven description of electronic correlation for valence and Rydberg states at CASSCF level is the cause for valence-Rydberg mixing [20] which affects electronic properties of valence states. Since the resulting Rydberg orbitals do not interact with the valence orbitals [22,18] we were able to apply the following scheme to resolve valence–Rydberg orbital mixing: An eight electrons/eight orbitals active space (i.e. CAS (8,8)) comprising all π-valence orbitals but the completely bonding π-orbital was constructed. The active space was then augmented with a Rydberg orbital of A''-symmetry (the Rydberg orbitals of A' symmetry are not considered further), thus forming CAS (8,9). Next a state-averaged calculation was performed where we made sure that at least one of the optimized states had Rydberg character. To prevent spurious orbital mixing the active space orbitals were localized using the Pipek procedure [23]. Finally, the pure Rydberg orbital was deleted from the orbital list. This procedure was repeated 24 times until all Rydberg orbitals were removed and, thus, removing also Rydberg excited states. Once the valence orbitals were clean from Rydberg contaminations the lowest occupied orbital was added to the active space, thus, forming the full π-valence active space of ten $\pi-$electrons in nine orbitals (i.e. CAS (10,9)) and the first 25 states with A'-symmetry were computed. To test the stability of the calculation the active space was systematically increased by four, eight and twelve additional extravalence virtual orbitals which were selected in consecutive order according to their energies (diagonal elements of the effective state averaged Fock matrix). To reduce computational cost only configurations with up to two excited electrons in these virtual orbitals were added to the list of configuration state functions. The resulting restricted active spaces (RAS) are labeled RAS $(0,0|10,9|2,4)$, RAS$(0,0|10,9|2,8)$ and RAS $(0,0|10,9|2,12)$, where each of the three restricted active subspaces are characterized by the maximum number of simultaneously excited electrons and the number of orbitals. The contributions of the remaining configurations were treated perturbationally with the multiconfigurational counterpart of the Møller–Plesset method denoted as CASPT2/RASPT2 [13]. All virtual orbitals were correlated in the perturbation procedure except for the 24 removed Rydberg orbitals with A''-symmetry. An IPEA shift of 0.0 a.u. [24] and an imaginary shift of 0.2 a.u. [25] were used. Transition dipole moments were calculated at SA-CASSCF/RASSCF level. The RICD approximation was used to speed up the calculation of two-electron integrals [26].

Aqueous solvation was described with the conductor-like polarizaable continuum model (C-PCM) technique [27,28], which describes the solvent as a polarizable dielectric continuum. The solute, which occupies a cavity defined as the envelope of spheres centered at non-hydrogen atoms, polarizes the dielectric medium and creates an electric field which, in turn, disturbs the solute. Convergence is achieved iteratively at the CASSCF level. Vertical electronic transitions were treated as non-equilibrium events. The solvent was allowed to equilibrate to the electronic density of the ground state. For excited states calculations only the fast electronic component of the electric field reflecting the instantaneously adapting electronic density of the solvent to the excited state density of the solute was allowed to equilibrate. The slow component reflecting solvent rearrangement was kept fixed. At CASPT2 level the reaction field was approximated as a static perturbation to the Hamiltonian. Default values were used for the dielectric constant of water ($\epsilon = 78.39$) and for the areas of the tesserae (0.4 Å^2).

All quantum–mechanical calculations were carried out with software package Molcas 7.7 [26].

Using the SS-CASPT2/RASPT2 energies and SA-CASSCF transitions dipole moments rephasing (denoted as KI), quasi-absorptive (denoted as PP) and double coherence (denoted as KIII) spectra were computed by the sum-over-states approach [29] with Spectron 2.7 [9]. The experimental set-ups use three pulses with wavevectors $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$ and a local oscillator with a wavevector $\mathbf{k}_4$ which satisfies the phase-matching conditions $\mathbf{k}_4 = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$ (KI), $\mathbf{k}_4 = \mathbf{k}_3$ (PP, satisfies both the rephasing KI and non-rephasing KII phase-matching conditions) and $\mathbf{k}_4 = \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3$ (KIII). A constant broadening of 200 cm^{-1} was used for all transitions. Calculations were preformed for the non-chiral *xxxx*, *xyxy* and *xyyx* pulse polarization sequences. The nonlinear signal depends parametrically on the three time delays t_1, t_2, t_3. The KI and PP spectra were calculated by 2D Fourier transformation with respect to t_1 and t_3 while setting t_2 to zero. In the case of KIII spectra 2D Fourier transformation was performed with respect to t_2 and t_3 while setting t_1 to zero. 2D spectra are plotted on a logarithmic scale. By convention bleach and stimulated emission contributions appear as negative (blue) signals, stimulated absorptions appear as positive (red) peaks for *xxxx* polarized spectra. This does not apply to cross-polarized pulse sequences.

3. Results and discussion

3.1. Absorption spectrum of indole

Indole has nine valence π-orbitals, five of which are occupied. The two highest occupied orbitals (HOMO-1 and HOMO) as well as the two lowest unoccupied orbitals (i.e. LUMO and LUMO + 1)

can be viewed as benzene-like orbitals (Fig. 1). For historical reasons the states associated with electronic transitions among these orbitals are named L_b, L_a, B_b and B_a following Platt's notation [30]. These four transitions are mainly responsible for the absorption spectrum at lower energies (i.e. below 6.5 eV) and have been extensively studied both experimentally and theoretically in vacuo and in solution. An overview is provided in Refs. [31,32]. Theoretical studies of the excited state spectrum up to the first ionization potential can be found in Ref. [19]. However, the electronic nature of the higher excited states is completely unknown.

Classification of higher lying excited states of polycyclic aromatic compounds beyond the $L_{a/b}$ and $B_{a/b}$ states is ambiguous due to the multiconfigurational nature of the excited states and their high spectral density which causes state mixing at different computational levels. Here we propose a classification of the singly excited manifold of indole and related aromatic compounds into bands based on the orbitals involved in the electronic transitions. Besides the HOMO and the HOMO-1 indole possesses three further bonding orbitals: a pyrrole-like orbital HOMO-2, another benzene-like orbital HOMO-3 which is completely bonding on the six-ring and a completely bonding pyrrole-like orbital HOMO-4 with large contribution from the nitrogen lone pair (Fig. 1). HOMO-2 and HOMO-3 have anti-bonding counterparts LUMO + 2 and LUMO + 3. The L_b and L_a transitions which involve the HOMO-1, HOMO, LUMO and LUMO + 1 orbitals form the lowest band, named L^+-band. The B_b and the B_a transitions are combined in the L^--band (Fig. 1, magenta). The + and − signs indicate that the two bands emerge through different linear combinations of the same transitions. Transitions from HOMO-2 into LUMO and LUMO + 1, which are degenerate to the transitions out of HOMO and HOMO-1 into LUMO + 2 form the A^+- and the A^--bands (Fig. 1, red), each one consisting of two transitions. Likewise, transitions from HOMO-3 into LUMO and LUMO + 1 which are degenerate to the transitions out of HOMO and HOMO-1 into LUMO + 3 are combined in the B^+- and the B^--bands (Fig. 1, blue). The HOMO-2 to LUMO + 2 transition is labeled *C*-band (Fig. 1, brown). The transitions from HOMO-2 into LUMO + 3 and from HOMO-3 into LUMO + 2 form the D^+- and the D^--bands (Fig. 1, green), each one consisting of a single transition. Finally, the HOMO-3 to LUMO + 3 transition is labeled *E*-band (Fig. 1, black). We did not find large CI-coefficients for configurations involving excitation out of the completely bonding orbital (HOMO-4) in the spectral region of interest. Therefore, the transitions out of HOMO-4 are not explicitly labeled. While the *L*- and *B*-bands are present in the UV-spectra of benzene and phenol [33] the *A*-, *C*- and *D*-bands, which involve the HOMO-2 and the LUMO + 2 belong only to the UV-spectrum of indole. Therefore, the proposed nomenclature is useful for qualitative assignment of the states and it allows recognizing transitions that are characteristic of the tryptophan, when compared to other protein chromophores absorbing in the NUV as benzene and phenol.

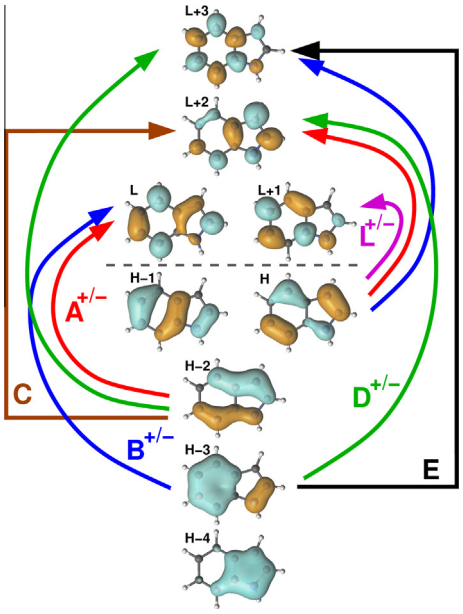

Fig. 1. List of the valence π-orbitals of indole together with different single electron transitions which determine the low energy part of the absorption spectrum of indole. Nearly degenerate transitions that are likely to mix are combined in groups denoted by a color code. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1 summarizes the electronic spectrum below 10 eV. The energy spectrum below the first ionization potential has ten excited states giving rise to five characteristic regions due to near degeneracy: 4.35 eV, 4.90 eV, 6.00 eV, 6.50 eV and 7.65 eV. In contrast to benzene and phenol indole has two characteristic transitions in the NUV. Our gas-phase calculations are in sound agreement with experimental findings and previous calculations in vacuo[19,31], slightly overestimating the absorption of L_a by ~0.10 eV. The L^+-band has a relatively weak oscillator strength (0.37 a.u. and 0.81 a.u.). Due to the large permanent dipole moment of 2.16 a.u. (compared to 0.60 a.u. for the ground state) of the L_a state we observe a red-shift of the ground state (GS) → L_a transition by ~0.10 eV in aqueous solution, even though less pronounced than the shift of 0.17 eV observed experimentally [34]. Due to the small energy gap between L_b and L_a their vibronic coupling is predicted to be very strong [34]. Hence, one can assume that both states are nearly simultaneously populated at sufficiently high excitation energies. *It is therefore intriguing to study if 2D spectroscopy would facilitate the state selective observation of excited state dynamics already in the monomer.* Open questions regarding the effective population transfer between the dominantly populated L_a state and the L_b and the time scale of the process, as well as regarding the decrease of the fluorescence quantum yield with increasing excitation energy [35] can be addressed.

The L^--, A^+-, A^-- and B^+-bands dominate the far-UV spectrum of indole below the ionization potential, giving rise to the intensive absorptions around 6.00 eV, 6.50 eV and 7.65 eV. Due to near degeneracy the bands mix strongly and entangle. The peak maxima are in agreement with experimental data [36]. It must be noted, though, that our calculations give reversed intensities for the absorption bands at 6.00 eV and 6.50 eV with respect to the experiment. In aqueous solution these states exhibit nonuniform shifts of up to ~0.10 eV. The A_2^- and the B_2^+ states (states 10, 11 in Table 1) lie at the ionization limit with solvatochromic shifts of 0.05–0.10 eV, thus, depending on the solvent those states might be shifted below or above it. Above the ionization potential the spectrum is characterized by a number of doubly excited states that emerge by promoting two electrons from HOMO and HOMO-1 to LUMO and LUMO + 1 (states 12, 15–18, 21, 22). They remain largely unaffected by aqueous solvation. In vacuo the *C*-band is found around 8.50 eV mixing with doubly excited configurations, while in aqueous solution it disentangles giving a clear signature at 8.40 eV. The B^--band (states 19, 20 in Table 1) is found around 9.70 eV, at significantly higher energies than its counterpart B^+. We found comparably large *B*-band splitting also in benzene and phenol [33]. States with clear *D*- or *E*-band signatures were not found in the investigated spectra region, but the D^+-band contributes strongly to states 15 and 21.

3.2. Dependence of the third-order non-linear response on electronic correlation

The L_b transition in indole has a weaker oscillator strength compared to the transition to the L_a state, an observation made also for

Table 1
Vertical excitation energies (in eV) in vacuo and aqueous solution, transition dipole moments (in a.u.) in vacuo out of the ground state (GS), the first (L_b) and the second (L_a) excited states calculated at RAS (0, 0|10, 9|2, 12)//SS-RASPT2 level.

	E (gas-phase)	E (water)	TDM (gas-phase) from			Main	
			GS	L_b	L_a	Conf.	Coeff.
2A' (L_1^+)	4.35 (4.37[a])	4.36 (4.31[a])	0.37	–	–	H-1 → L	−0.67
						H → L + 1	0.51
3A' (L_2^+)	4.91 (4.77[a])	4.82 (4.59[a])	0.81	0.21	–	H → L	−0.77
4A' (A_1^+)	5.96	5.98	0.36	0.10	0.32	H-1 → L + 2	−0.37
						H → L + 2	−0.34
						H-2 → L + 1	0.33
5A' ($L_1^- + A_2^+$)	6.00 (6.02[a])	5.95	0.95	0.09	0.72	H → L + 1	0.45
						H-2 → L	0.34
						H-1 → L + 2	0.32
						H → L + 2	−0.31
6A' ($L_2^- + Bs_1^+$)	6.45 (6.35[a])	6.49	1.41	0.45	0.16	H-1 → L + 1	−0.58
						H-3 → L	−0.30
						H → L	0.30
7A' ($L_1^- - A_2^+$)	6.46	6.51	1.74	0.61	0.45	H → L + 1	0.42
						H-2 → L	−0.42
						H-1 → L	0.38
						H-1 → L + 2	−0.33
8A' ($B_1^+ - L_2^-$)	6.48	6.42	0.81	0.15	0.41	H-3 → L	−0.45
						H-1 → L + 1	0.38
						H → L + 3	−0.20
9A' (A_1^-)	6.74	6.70	1.42	0.26	0.23	H → L + 2	−0.50
						H-2 → L	−0.44
10A' (A_2^-)	7.65	7.73	0.66	0.81	0.32	H-2 → L + 1	−0.48
						H-1 → L + 2	−0.38
11A' (B_2^+)	7.67	7.64	0.98	0.15	0.25	H-3 → H + 1	0.41
						H-2 → L + 1	−0.35
						H-1 → L + 3	0.30
12A'	8.15	8.16	1.03	0.29	0.82	H ⇒ L	−0.40
13A' (C)	8.47	8.40	0.81	0.03	0.58	H-2 → L + 2	0.41
						H-3 → L + 2	−0.31
14A' (C)	8.49	8.57	1.37	0.06	0.87	H-2 → L + 2	−0.49
15A'	8.83	8.87	0.45	0.04	0.32	H ⇒ L,L + 1	−0.36
						H-3 → L + 2	0.28
16A'	8.92	8.95	0.36	0.37	0.32	H-1 ⇒ L	0.33
17A'	9.10	9.06	0.26	0.26	1.14	H ⇒ L,L + 1	0.31
						H-1 ⇒ L	−0.30
18A'	9.24	9.23	0.54	0.34	0.43	H-1,H ⇒ L	−0.35
						H ⇒ L,L + 1	0.28
19A' (B_1^-)	9.62	9.63	0.27	0.07	0.38	H-4 → L	−0.36
						H → L + 3	−0.29
20A' (B_2^-)	9.78	9.80	0.36	0.43	0.70	H-1 → L + 3	−0.28
						H-1,H ⇒ L,L + 1	−0.27
						H-3 → L + 1	0.21
21A'	9.81	9.81	0.23	0.34	0.41	H-3 → L + 2	0.26
						H ⇒ L,L + 2	−0.24
						H-4 → L + 1	0.24
22A'	9.94	9.92	0.15	0.14	1.31	H-1 ⇒ L,L + 1	−0.44
						H ⇒ L,L + 1	−0.25

[a] Experimental values (in brackets) adopted from Ref. [19].

benzene and phenol [37]. *To achieve good resolution of both states it is essential to use a narrow-band pump pulse pair centered closer to the frequency of the L_b transition in order to enhance it.* On the other hand broadband pulses are desired to probe large spectral windows and, thus, resolve characteristic signals. In this work bandwidths with a full width at half maximum (FWHM) of 1466 cm^{-1} (corresponding to a Fourier limited pulse of 10.0 fs) were used for pumping, while probing was performed with pulses with FWHM of 5864 cm^{-1}. We note, however, that the spectra presented here can be obtained also with narrow-band probe pulses tunable in the range of interest. Pump pules were centered between 36,000 cm^{-1} and 37,000 cm^{-1}, depending on the level of calculation (Fig. 2). Probing was performed in two spectral windows: NUV and Vis. Probing in the NUV resolves correlated transitions in both monomers and coupled aggregates. For indole we probe the spectral region between 32,000 cm^{-1} and 42,000 cm^{-1} to resolve the correlation between the L_b and L_a states of the monomer. Therefore, the probe pulse pair was centered in the same region as the pump pulse pair (i. e. between 36,000 cm^{-1} and 37,000 cm^{-1}, one-color regime, Fig. 2). Probing in the Vis covers the spectral region below the ionization potential and permits the collection of background-free signals. *This is particularly interesting for the sake of characterizing charge transfer states in coupled chromophore aggregates*, since inter-chromophore interactions are relatively weak and require regions free of background noise and overlapping intensive local absorptions [33]. As we do not know *a priori* where the charge transfer transitions are likely to appear we probe a broad spectral window between 10,000 cm^{-1} and 26,000 cm^{-1} covering the near-IR to Vis region (Fig. 2). The probe pulse is centered at 18,000 cm^{-1} (i.e. two-color regime).

Fig. 3a–d shows a comparison of gas-phase 2D NUV-pump NUV-probe KI spectra of the indole monomer for different active space sizes. All pulses have the same polarization (i.e. *xxxx*). The spectra are dominated by the diagonal ($\Omega_{1/3} = 35,000\ cm^{-1}$ and

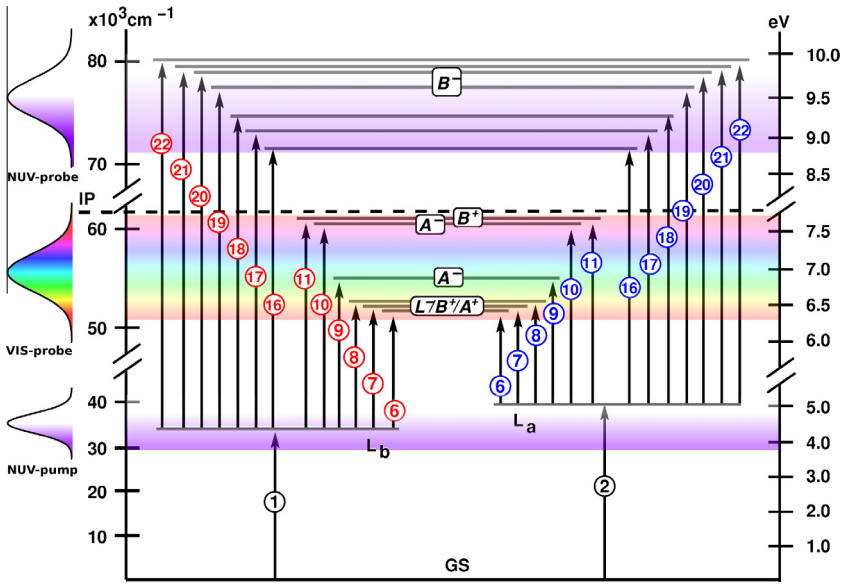

Fig. 2. Level scheme of indole monomer including the higher excited states that can be probed out of the L_b and L_a states by using broadband pulses centered in the Vis and NUV spectral windows. Signals associated with transitions from the L_b are shown in red, signals associated with transitions from the L_a are shown in blue. The numbering corresponds to the RAS (0, 0|10, 9|2, 12)//SS-RASPT2 state ordering in vacuo (Table 1, 2nd column). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

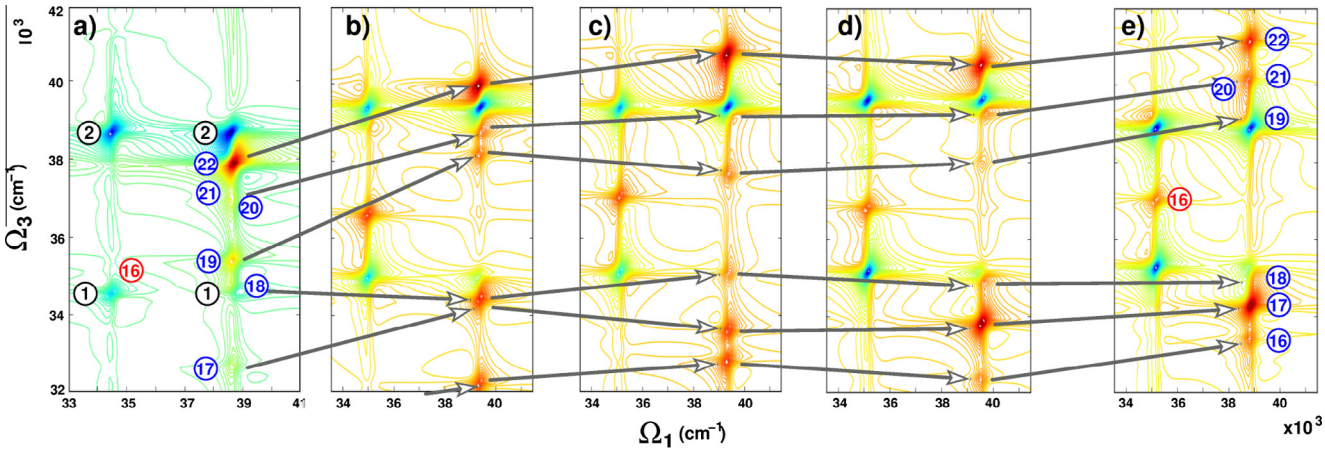

Fig. 3. (a–d) Comparison of gas-phase 2D pump-NUV probe-NUV KI spectra for different active space sizes: CAS (10,9) (a), RAS (0, 0|2, 4) (b), RAS (0, 0|2, 8) (c), RAS (0, 0|2, 12) (d). (e) 2D pump-NUV probe-NUV spectrum with RAS (0, 0|2, 12) in aqueous solution. Labeling according to Fig. 2.

$\Omega_{1/3} = 39,000\ \text{cm}^{-1}$) and off-diagonal ($\Omega_1 = 35,000\ \text{cm}^{-1}$/ $\Omega_3 = 39,000\ \text{cm}^{-1}$) bleach (negative, blue) signals of the L^+-band. The diagonal signals consist of two contributions, ground state bleach and stimulated emission. The off-diagonal bleach signals appear for coupled transitions when a coherence between the ground state and a state from the singly excited manifold (here the L^+-band) created by the interaction of the system with the first pump pulse, is destroyed by the subsequent interaction with the second pump pulse, thereby allowing for another state from the singly excited manifold to be probed. For non-correlated oscillators these signals vanish due to exact cancellation of quantum pathways as given by different Feynman diagrams. The clear off-diagonal bleach signals demonstrate that the L_b and the L_a transitions are strongly correlated. *The off-diagonal bleach represent a characteristic fingerprint of the NUV 2D spectrum of indole* as the corresponding L_b/L_a couplings in benzene and phenol lie in the far-UV. The signals reflect intrinsic properties of the indole chromophore and are present in the monomer as well as in aggregates.

Several positive peaks corresponding to stimulated absorption from the L^+-band are present in the spectra with the most prominent ones being the transitions from L_b and L_a to states 16, 17 and 22. Table 1 shows that these are doubly excited states accessible out of L^+-band through an excitation of a second electron out of the HOMO or the HOMO-1 into LUMO or LUMO + 1. The absorption peak 18 cancels out the off-diagonal bleach at $\Omega_1 = 39,000\ \text{cm}^{-1}$/ $\Omega_3 = 35,000\ \text{cm}^{-1}$ and both signals no longer be clearly resolved.

The comparison of the different CAS/RAS schemes demonstrates the sensitivity of 2D electronic spectra for electronic correlation. A clear trend of blue shifting absorption peaks (e.g. peaks 15, 18, 19, 21) relatively to the bleach signals by approx. 2000 cm^{-1} (i.e. along Ω_3) is observed in Fig. 3. On absolute scale the shift is more pronounced due to the simultaneous blue shift of the L^+-band by approx. 1000 cm^{-1} (i.e. along Ω_1) and it is approx. 4000 cm^{-1}. *This demonstrates that valence active space calculations do not provide accurate energies of high lying excited states to reliably simulate 2D experiments.* While generally low lying states are described with

satisfying accuracy the dynamic correlation via the perturbation treatment is overestimated for higher lying states. Our calculations demonstrate that this problem can be overcome by including sufficient number of extravalence virtual orbitals in the active space. As the computational effort scales exponentially with the active space size the restricted active space (RAS) modification of the active space technique emerges as the natural methods of choice. We chose to include configurations with up to two excited electrons in the extravalence orbitals in the list of configuration state functions. The contributions of the remaining configurations were treated perturbationally. Fig. 3 demonstrates that the most significant change is observed when the lowest four extravalence orbitals are included in the active space (Fig. 3a and b). Adding more orbitals fine-tunes the positions of the peaks. The spectra stabilize with eight additional orbitals in the active space, while no significant effects are observed by adding more orbitals. Notably, the RASSCF state order is mostly preserved at RASPT2 level with only a few outliers which exhibit stronger stabilization compared to the states in their energetic vicinity. We chose RAS $(0,0|10,9|2,12)$ as the reference active space for obtaining spectra in aqueous solution.

In solution a pronounced blue-shift is observed for the excited state absorptions, correlated with the GS → L_a transition, while the absorption peaks along the L_b band do not shift (Fig. 3e). The reason for this behavior lies in the aforementioned solvatochromic shift of the L_a state. Considering the observation that the doubly excited manifold is not affected by the solvent (Table 1) the red shift of the L_a transition implies a blue shift for the transitions to the higher lying states.

Fig. 4) shows a study of the active space dependence of the pump-NUV probe-Vis 2D spectrum of indole in the range between 10,000 cm^{-1} and 26,000 cm^{-1}. In contrast to the pump-NUV probe-NUV spectrum only some transitions blue-shift significantly in the visible (e.g. peaks 6, 10, 11 in Fig. 4), while other transitions are already well described at CAS (10,9)/CASPT2 level (e.g. peaks 7, 8, 9 in Fig. 4). Latter exhibit only small fluctuations upon enlarging the active space size. Again, the largest shifts are observed when the active space is increased by four additional virtual orbitals (Fig. 4b), while the spectra stabilize with eight to twelve additional orbitals. *The existence of a broad absorption-free window between 18,000 cm^{-1} and 26,000 cm^{-1} when probing from the L_b state (corresponds to the region between 14,000 cm^{-1} and 22,000 cm^{-1} when probing from the L_a state) shows an exciting promise as a candidate for tracking charge transfer states in aggregates* [33]. Importantly, this window also persists in aqueous solution (Fig. 4e). It should be noted that the Vis probing region has absorptions, about an order of magnitude weaker than in the NUV window. We anticipate that this will help detecting weak charge transfer signals in coupled aggregates.

3.3. Comparison of different non-linear spectroscopy techniques

Detecting the third-order nonlinear system response for different phase matching conditions allows to select particular Feynman diagrams [38,39]. Fig. 5 shows the quasi-absorptive (PP, Fig. 5a) and double coherence (KIII, Fig. 5b) 2D pump-NUV probe-NUV spectra of solvated indole. The PP spectrum was generated with the same pulse parameters as for the rephasing signal KI (Fig. 3e). Quasi-absorptive spectra exploit the fact that phase variations which broaden the rephasing (KI) and non-rephasing (KII) spectra cancel when KI and KII spectra are added, resulting in clean spectra with highest possible resolution. This can be achieved by either collecting both the KI and KII signals separately (implies two measurements under the same experimental conditions) and adding them *posteriori* or directly measuring the quasi-absorptive signal using a pump–probe pulse pair oriented collinear [39]. Although the addition does not yield a perfectly absorptive spectrum in the case of coupled oscillators [39] PP spectra enhance the spectral resolution.

The KIII spectrum was generated with the following pulse parameters: the first pulse was centered at 35,500 cm^{-1} with a FWHM of 1466 cm^{-1}, the other three pulses were centered at 37,500 cm^{-1} with a FWHM of 2932 cm^{-1}. For KIII spectra interactions of the system with the first two pulses create a coherence between the ground state and the doubly excited manifold. The third pulse correlates this coherence to coherence between the ground state and the singly excited manifold (diagram A, inset of Fig. 5c) or to coherence between the singly and doubly excited manifolds (diagram B, inset of Fig. 5d). For oscillators which are harmonic (i.e. doubly excited states $|f\rangle$ absorb at twice the energy of the singly excited state $|e\rangle$) and uncorrelated (i.e. common doubly excited states $|f\rangle$ absorb at the sum of the singly excited states $|e\rangle$ and $|e'\rangle$) the signal vanishes as both diagrams A and B cancel exactly [40]. Thus, KIII spectra are a suitable for resolving local anharmonic shifts, as well as correlated transitions. In the Ω_2 vs. Ω_3 representation states common to both L_b and L_a characterize through a set of four peaks along Ω_3 for one Ω_2 value: two negative peaks at the transition frequencies of L_b (e.g. peak 16(II), Fig. 5b) and L_a (e.g. peak 16(IV), Fig. 5b) and two positive signals, equally red- or blue-shifted with respect to the negative peaks (e.g. red-shifted peaks 16(I) and 16(III), Fig. 5b). The shift δ resembles the

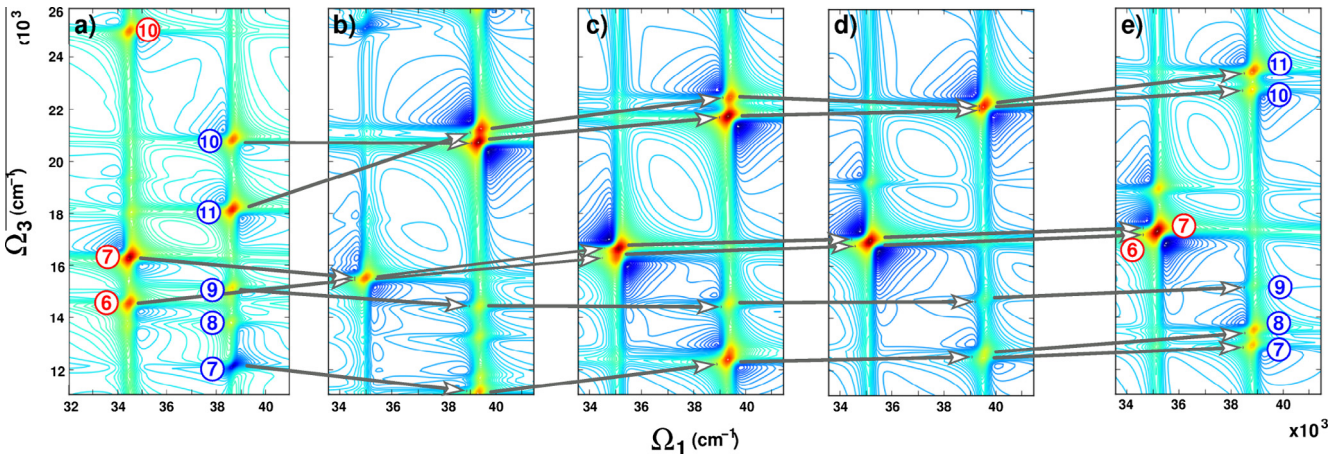

Fig. 4. (a–d) Comparison of the 2D pump-NUV probe-Vis spectra for different active space sizes obtained in gas-phase calculations: CAS (10,9) (a), RAS $(0,0|2,4)$ (b), RAS $(0,0|2,8)$ (c), RAS $(0,0|2,12)$ (d). (e) 2D pump-NUV probe-NUV spectrum with RAS $(0,0|2,12)$ in aqueous solution. Labeling according to Fig. 2.

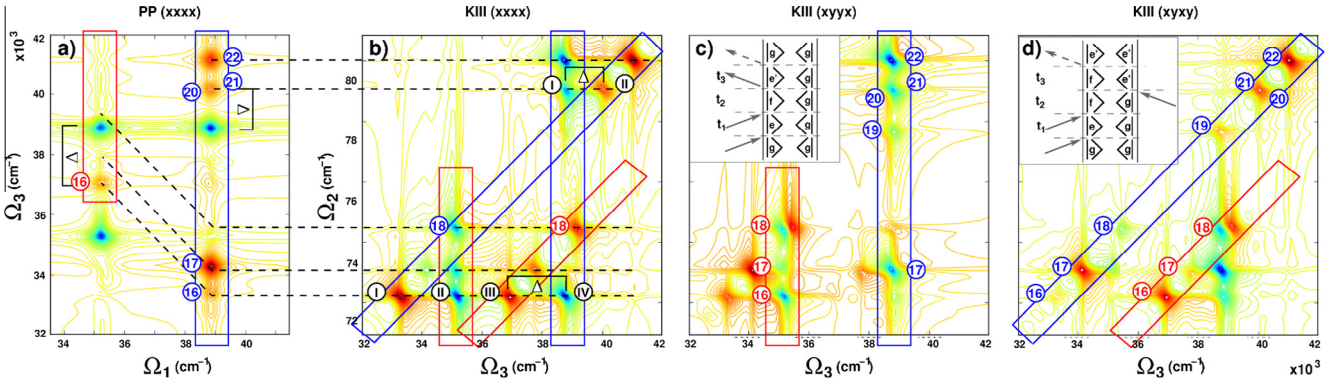

Fig. 5. Quasi-absorptive PP (a) and double coherence KIII (b–d) 2D pump-NUV probe-NUV spectra for indole in aqueous solution. For KIII cross-polarized spectra *xyyx* (c) and *xyxy* (d), which enhance particular Feynman diagrams (shown as insets) are shown. Labeling according to Fig. 2.

coupling strength. Excitations local to either L_b or L_a exhibit only two signals (e.g. peaks 20, 21 in Fig. 5b), one negative (e.g. 20, 21(I) in Fig. 5b) and one positive (e.g. 20, 21(II) in Fig. 5b) and the splitting δ reflects the anharmonicity shift.

The diagonal and off-diagonal bleaches dominating the KI and PP spectra (peaks 1 and 2 in Fig. 5a) do not appear in the KIII spectrum as they cannot be represented neither through diagram A nor through diagram B. Thus, spectral features hidden in the KI and PP spectra reveal (peaks 18, 19 in Fig. 5b–d). Regarding excited state absorptions, the KIII spectra contain the same information as the KI (Fig. 3e), KII or the PP (Fig. 5a) spectra. The difference is in the way of representing the information. In KI, KII and PP spectra transitions to states common to L_b and L_a appear red shifted along Ω_3 (e.g. peak 16 in Fig. 5a) with shift equal to the energy difference between L_b and L_a. In KIII spectra peaks associated with transitions to common states are ordered horizontally (peaks 16(I–IV)) and signals appear in pairs of positive and negative peaks. Therefore, KIII spectra contain about twice more peaks than KI spectra. The spectral congestion can make analysis of the spectra cumbersome. A remedy is to use cross-polarized pulse sequences *xyxy* or *xyyx*. These sequences make use of the fact that transitions excited with cross-polarized pulses result in weaker signals than transitions excited with equally polarized pulses. Exemplary, diagram A represents the up and down climbing of the energy ladder ($|g\rangle \rightarrow |e\rangle \rightarrow |f\rangle \rightarrow |e'\rangle \rightarrow |g\rangle$). For $|e'\rangle \equiv |e\rangle$ the polarization sequence *xyyx* uses *x*-polarized pulses to excite the identical $|g\rangle \rightarrow |e\rangle$ and $|e\rangle \rightarrow |g\rangle$ transitions and *y*-polarized pulses to excite the identical $|e\rangle \rightarrow |f\rangle$ and $|f\rangle \rightarrow |e\rangle$ transitions. The polarization sequence *xyxy* would use cross-polarized pulses for the aforementioned identical transitions, yielding generally a weaker signal. As shown in Fig. 5c and d cross polarization schemes can select signals belonging to different Feynman diagrams, separating groups of signals that are merged in the *xxxx* spectrum.

Cross-polarized pulse sequences find application also in KI and PP spectroscopy [41]. Absorptions from the L^+-band to common states from the doubly excited manifold, as well as off-diagonal bleach signals can be selectively enhanced by recording the difference spectrum for two polarization conditions (e.g. *xyxy–xyyx*). Diagonal bleaches, stimulated emission and local absorptions are thereby effectively removed. Only pathways for which the third and fourth system-laser pulse interactions promote different excitations become visible. The KI and PP difference spectra in the NUV

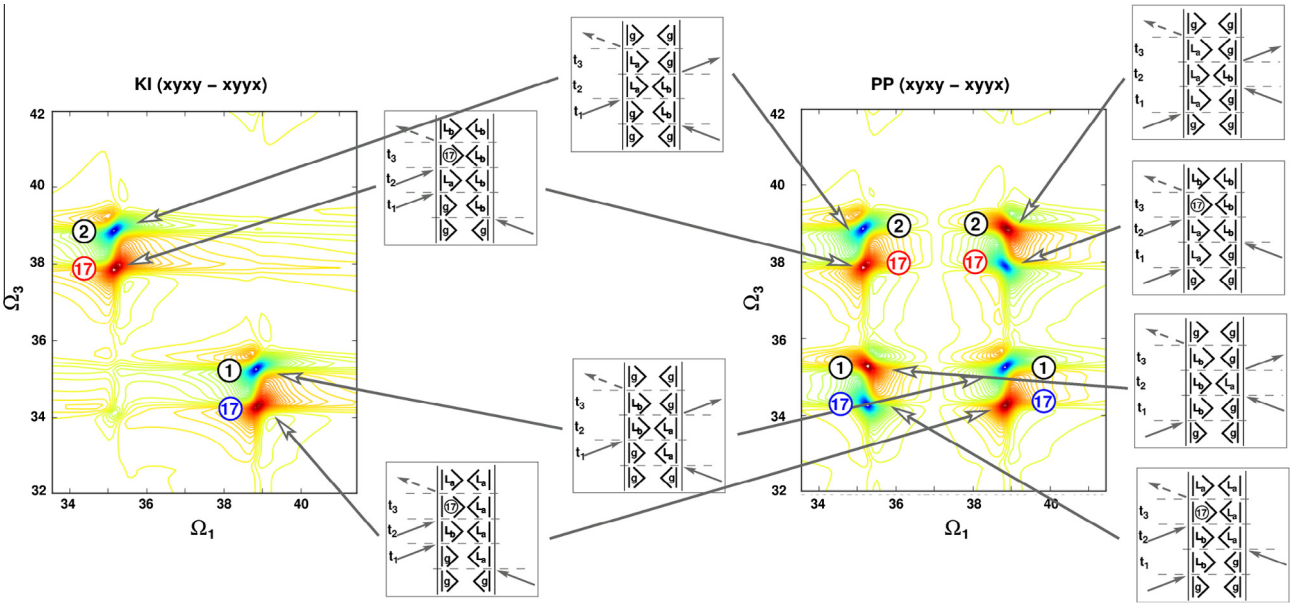

Fig. 6. *xyxy – xyyx* difference rephasing (a) and quasi-absorptive (b) 2D pump-NUV probe-NUV spectra for indole in aqueous solution. Explicit Feynman diagrams are for each peak are shown. Labeling according to Fig. 2.

probing region are shown in Fig. 6 together with the corresponding Feynman diagrams. The off-diagonal bleach signals 1 and 2 as well as the absorption into the common state 17 are the only contributions to the KI spectrum. The PP spectrum contains twice more signals because replica of the off-diagonal peaks appear close to the diagonal. These signals do not provide additional information and can complicate the interpretation of the spectra, especially in complex systems. It strikes to our attention that the common peak 16 is not resolved in these spectra, although the KIII (Fig. 5b) clearly indicates that it is accessible from both L_b and L_a. A look at the transitions dipole moments from the L_b and the L_a states to state 16 reveals that their vectors are parallel and of similar magnitude. The difference is too small for the signals to be discriminated via pulse polarized spectroscopy.

4. Conclusion

The UV spectrum of indole, the aromatic side chain of the amino acid tryptophan, is characterized by using multiconfigurational methods, showing two bright low-energy transitions in the NUV, in contrast to benzene and phenol, the chromophores of the amino acids phenylalanine and tyrosine, which possess only one transition. A nomenclature for electronic transitions is proposed, based on the orbitals involved in the transitions from the ground state, making clear distinction between the transitions common to all aromatic amino acids (bands *L* and *B*) and those characteristic for tryptophan (bands *A* and *C*). We report simulated 2DES spectra of indole in gas-phase and in aqueous solution using different experimental setups, including rephasing KI signal in the one- and two-color regimes (with pump pulses in the NUV and probe pulses in the NUV and the Vis, respectively), and quasi-absorptive (PP) and double quantum coherence (KIII) signals. The gas-phase 2D NUV-pump NUV-probe KI spectrum has been simulated using different active space sizes, indicating that inclusion of extravalence virtual orbitals is required to obtain accurate energies of high lying excited states and reliable 2DES spectra. The one-color KI spectra in vacuo and in solution are dominated by the diagonal and off-diagonal bleach signals of the L_b and L_a transitions, with well-resolved off-diagonal peak at $\Omega_1 = 35,000\ \mathrm{cm}^{-1}/\Omega_3 = 39,000\ \mathrm{cm}^{-1}$ and its symmetric counterpart at $\Omega_1 = 39,000\ \mathrm{cm}^{-1}/\Omega_3 = 35,000\ \mathrm{cm}^{-1}$ being covered by nearby excited state absorptions peaks. The proximity of the L_b and L_a states determines the presence of these off-diagonal signals in the 2D-NUV spectrum and it clearly differentiates indole from benzene and phenol. However, electronic states above the ionization potential of the chromophore are involved in the one-color 2D-NUV spectra with the possible drawback of underlying background signal. A background-free region bellow the ionization potential can be accessed in 2D-NUV/Vis two-color experiments. Characteristic transitions are resolved with excited states absorptions having generally weaker intensity than NUV absorption. A background-free region between 53,000 and 61,000 cm^{-1}, accessible by probing in the Vis, shows no absorption in the monomer and we anticipate that it can be used to resolve charge transfer states in coupled chromophore aggregates.

This work also provides fundamental information for design of the 2DES experiments of indole, showing how various 2D spectroscopy techniques can be used to achieve high spectral resolution or extract particular information from the spectra. We show that quasi-absorptive spectra allow collecting well-focused spectra without phase twists and that double coherence spectra are well-suited to extract anharmonicities. Difference spectra of cross-polarized pulse sequences (e.g. *xyxy* – *xyyx*) are very appealing for resolving congested spectra since they extract transitions to common states. The proposed cross-analysing of different 2DES spectra will provide unprecedented level of detail.

Acknowledgment

A.N. and M.G. wish to thank Dr. Francesco Aquilante for his help in performing the high level calculations and Anna Stoyanova, founder of the "Hamiltonian of Design" studio (http://hamiltonianofdesign.wordpress.com) for creating the TOC graphic. M.G. acknowledges support by the European Research Council Advanced Grant STRATUS (ERC-2011-AdG No. 291198). S.M. gratefully acknowledges the support of the National Science Foundation (Grant No. CHE-1058791) and the National Institute of Health (Grant R01- GM-59230).

References

[1] H. Sindrasky, Tryptophan: Biochemical and Health Implications, CRC Series in Modern Nutrition, CRC Press, Boca Raton, Florida, USA.
[2] A.L. Mallam, S.E. Jackson, Molecular Biology of Protein Folding, Part B, Progress in Molecular Biology and Translational Science, vol. 84, Elsevier, London, UK, 2008.
[3] W.C. Johnson, Secondary structure of proteins through circular-dichroism spectroscopy, Annu. Rev. Biophys. Bio. 17 (1988) 145–166.
[4] C. Kolano, J. Helbing, M. Kozinski, W. Sander, P. Hamm, Watching hydrogen-bond dynamics in a β-turn by transient two-dimensional infrared spectroscopy, Nature 444 (2006) 469–472.
[5] Y.C. Cheng, G.R. Fleming, Dynamics of light harvesting in photosynthesis, Annu. Rev. Phys. Chem. 60 (2009) 241–262.
[6] B.A. West, A.M. Moran, Two-dimensional electronic spectroscopy in the ultraviolet wavelength range, J. Phys. Chem. Lett. 3 (18) (2012) 2575–2581.
[7] D.B. Turner, P.C. Arpin, S.D. McClure, D.J. Ulness, G.D. Scholes, Coherent multidimensional optical spectra measured using incoherent light, Nature Commun. 4.
[8] M. Cho, Two-Dimensional Optical Spectroscopy, CRC Press, Boca Raton, Florida, USA, 2009.
[9] D. Abramavicius, B. Palmieri, D.V. Voronine, F. Sanda, S. Mukamel, Coherent multidimensional optical spectroscopy of excitons in molecular aggregates; quasiparticle versus supermolecule perspectives, Chem. Rev. 109 (6) (2009) 2350–2408.
[10] S. Mukamel, D. Healion, Y. Zhang, J.D. Biggs, Multidimensional attosecond resonant X-ray spectroscopy of molecules: Lessons from the optical regime, Annu. Rev. Phys. Chem. 64 (2013) 101–127.
[11] I. Rivalta, A. Nenov, G. Cerullo, S. Mukamel, M. Garavelli, Ab initio simulations of two-dimensional electronic spectra: the sos//qm/mm approach, Int. J. Quant. Chem. 114 (2) (2014) 85–93.
[12] B.O. Roos, Ab Initio Methods in Quantum Chemistry: Part II, Wiley, Chicester, UK, 1987.
[13] K. Andersson, P.A. Malmqvist, B.O. Roos, A.J. Sadlej, K. Wolinski, 2nd-order perturbation-theory with a casscf reference function, J. Phys. Chem. 94 (14) (1990) 5483–5488.
[14] T.Q. Liu, P.R. Callis, B.H. Hesp, M. de Groot, W.J. Buma, J. Broos, Ionization potentials of fluoroindoles and the origin of nonexponential tryptophan fluorescence decay in proteins, J. Am. Chem. Soc. 127 (11) (2005) 4104–4113.
[15] P.O. Widmark, P.A. Malmqvist, B.O. Roos, Density matrix averaged atomic natural orbitals (ano) basis sets for correlated molecular wave functions. 1. 1st row atoms, Theor. Chim. Acta 77 (5) (1990) 291–306.
[16] M.P. Fulscher, L. Serrano Andres, B.O. Roos, A theoretical study of the electronic spectra of adenine and guanine, J. Am. Chem. Soc. 119 (26) (1997) 6168–6176.
[17] L. Serrano Andres, M.P. Fulscher, Theoretical study of the electronic spectroscopy of peptides.1. The peptidic bond: primary, secondary, and tertiary amides, J. Am. Chem. Soc. 118 (48) (1996) 12190–12199.
[18] J. Lorentzon, M.P. Fülscher, B.O. Roos, Theoretical study of the electronic spectra of uracil and thymine, J. Am. Chem. Soc. 117 (36) (1995) 9265–9273.
[19] L. Serrano Andres, B.O. Roos, Theoretical study of the absorption and emission spectra of indole in the gas phase and in a solvent, J. Am. Chem. Soc. 118 (1) (1996) 185–195.
[20] B.O. Roos, M.P. Fülscher, Malmqvist, P.-Å., Merchan, L.M. Serrano-Andres, Theoretical studies of electronic spectra of organic molecules, in: S.R. Langhoff (Ed.), Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.
[21] K. Kaufmann, W. Baumeister, M. Jungen, Universal Gaussian-basis sets for an optimum representation of rydberg and continuum wavefunctions, J. Phys. B-At. Mol. Opt. 22 (14) (1989) 2223–2240.
[22] L. Serranoandres, M. Merchan, I. Nebotgil, R. Lindh, B.O. Roos, Towards an accurate molecular orbital theory for excited states – ethene, butadiene and hexatriene, J. Chem. Phys. 98 (4) (1993) 3151–3162.
[23] J. Pipek, P.G. Mezey, A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions, J. Chem. Phys. 90 (9) (1989) 4916–4926.
[24] G. Ghigo, B.O. Roos, P.A. Malmqvist, A modified definition of the zeroth-order hamiltonian in multiconfigurational perturbation theory (caspt2), Chem. Phys. Lett. 396 (1–3) (2004) 142–149.
[25] N. Forsberg, P.A. Malmqvist, Multiconfiguration perturbation theory with imaginary level shift, Chem. Phys. Lett. 274 (1-3) (1997) 196–204.

[26] F. Aquilante, L. De Vico, N. Ferre, G. Ghigo, P.A. Malmqvist, P. Neograd y, T.B. Pedersen, M. Pitonak, M. Reiher, B.O. Roos, L. Serrano-Andres, M. Urban, V. Veryazov, R. Lindh, Software news and update molcas 7: the next generation, J. Comput. Chem. 31 (1) (2010) 224–247.
[27] V. Barone, M. Cossi, Quantum calculation of molecular energies and energy gradients in solution by a conductor ssolvent model, J. Phys. Chem. A 102 (11) (1998) 1995–2001.
[28] M. Cossi, N. Rega, G. Scalmani, V. Barone, Polarizable dielectric model of solvation with inclusion of charge penetration effects, J. Chem. Phys. 114 (13) (2001) 5691–5701.
[29] G.H. Chen, S. Mukamel, D. Beljonne, J.L. Bredas, The coupled electronic oscillators vs the sum-over-states pictures for the optical response of octatetraene, J. Chem. Phys. 104 (14) (1996) 5406–5414.
[30] J.R. Platt, Classification of spectra of cata-condensed hydrocarbons, J. Chem. Phys. 17 (5) (1949) 484.
[31] A. Giussani, M. Merchan, D. Roca-Sanjuan, R. Lindh, Essential on the photophysics and photochemistry of the indole chromophore by using a totally unconstrained theoretical approach, J. Chem. Theory Comput. 7 (12) (2011) 4088–4096.
[32] D.M. Rogers, J.D. Hirst, Ab initio study of aromatic side chains of amino acids in gas phase and solution, J. Phys. Chem. A 107 (50) (2003) 11191–11200.
[33] A. Nenov, I. Rivalta, G. Cerullo, S. Mukamel, M. Garavelli, Disentangling peptide configurations via two-dimensional electronic spectroscopy: ab initio simulations beyond the frenkel exciton hamiltonian, J. Phys. Chem. Lett. 5 (4) (2014) 767–771.
[34] C. Brand, J. Kuepper, D.W. Pratt, W.L. Meerts, D. Kruegler, J. Tatchen, M. Schmitt, Vibronic coupling in indole: I. Theoretical description of the (1)l(a)-(1)l(b) interaction and the electronic spectrum, Phys. Chem. Chem. Phys. 12 (19) (2010) 4968–4979.
[35] N. Glasser, H. Lami, Temperature-dependence of the nonradiative decay of indoles in solution, J. Mol. Struct. 142 (1986) 193–196.
[36] P. Ilich, S.S. Sedarous, Indole in argon matrix: the near uv spectra, Spectrosc. Lett. 27 (8) (1994) 1023–1039.
[37] J. Lorentzon, P.A. Malkmqvist, M.P. Fülscher, B.O. Roos, A caspt2 study of the valence and lowest rydberg electronic states of benzene and phenol, Theor. Chim. Acta 91 (1994) 91–108.
[38] S. Mukamel, Principles of Nonlinear Optics and Spectroscopy, Oxford University Press, Oxford, UK, 1995.
[39] P. Hamm, M. Zanni, Concepts and Methods of 2D Infrared Spectroscopy, Cambridge University Press, Cambridge, UK, 2011.
[40] Z. Li, D. Abramavicius, S. Mukamel, Probing electron correlations in molecules by two-dimensional coherent optical spectroscopy, J. Am. Chem. Soc. 130 (2008) 3509–3515.
[41] B.P. Fingerhut, S. Mukamel, Resolving the electron transfer kinetics in the bacterial reaction center by pulse polarized 2-d photon echo spectroscopy, J. Phys. Chem. Lett. 3 (13) (2012) 1798–1805.